



Design knowledge extraction from scenario-based databases using associative search engine for FR-induced decisions

M. Nakao (2)^{a,*}, K. Tsuchiya^a, K. Iino^b

^a Department of Engineering Synthesis, School of Engineering, The University of Tokyo, Tokyo, Japan

^b SYDROSE LP, CA, USA

ARTICLE INFO

Keywords:

Design method
Knowledge management
Decision making

ABSTRACT

The authors set the functional requirement of “reduce your own risk” and ran sessions of associative search to extract useful knowledge from databases. Ninety engineers found the most analogous knowledge using the associative search engine “GETA/IMAGINE” on scenario-based databases “Failure Knowledge Database” and “100 Scenarios of Failure.” Above 60% of their risk concern cases successfully reached the most analogous accident cases or failure scenarios from either database in about 10 min. Associative search can aid the designer in selecting design solutions for the functional domain.

© 2008 CIRP.

1. Introduction

Designers can use past knowledge in designing products. This knowledge application is effective in all stages of design. Fig. 1 shows the design process defined by Suh in Axiomatic Design [1]. The designer, as his business strategy, plans the customer attribute (CA) and analyzes the CA to define his design specification, the functional requirements (FR). He further selects the design parameters (DP) to satisfy the FR and determines the process variables (PV) to realize the DP. The designer decides CA, FR, DP, and PV in this order and each decision takes special knowledge.

Search technology has made rapid advancement with the recent development of information technology and the Internet. Especially in the 1990s, CAD tools were developed to determine the PVs like dimensions, material, and cutting tools automatically with algorithms from past knowledge [2,3].

Knowledge can, similarly, be applied in the upstream of Fig. 1 [4]. But deciding general DP like mechanism, motion, or fabrication process is different from setting specific PV and is difficult to apply knowledge-based methods. To form the DP, the designer first has to express FR in words. The FR, however, often involve organizational and tacit knowledge in addition to technical knowledge. Lu and Tseng pointed out that the designer should also be familiar with the mechanisms of decision making and negotiation within his organization [5,6].

FR lies in the functional domain of Fig. 1, and thus is expressed in subjective phrases, not in objective topography. Typical AND searches with subjective, non-normalized, non-engineering phrases as keywords give outputs that are either too large or empty, and the search sessions often take too long. In our study, we conducted OR searches by automatically selecting keywords from sentences that describe the FR as Fig. 2 shows. The number of hits

from the search becomes large; however, we can calculate the similarity score at the same time and output only those with high score. The selected keywords like “loosened-bolt, vibrating, fatigue” resemble the game “word association” and thus this OR search is called “associative search.”

The purpose of our study is demonstrate the effect of FR-induced decision making by extracting useful knowledge from knowledge databases with natural language processing techniques. The study started by assigning the FR of “reduce your own risk” to the designer, had him apply associative search against knowledge databases of accident cases, and evaluated whether the search sessions found useful failure knowledge for the designer. Studies for risk management have been made by extracting knowledge from accident databases [7,8]. Our study quantitatively evaluated the effect and issues of our knowledge extraction method through experiments.

2. Experimental method

2.1. Associative search engine

The authors selected “GETA (Generic Engine for Transposable Association, <http://geta.ex.nii.ac.jp>) [9]” for executing our associative searches. GETA, designed by Nishida et al. is a piece of open source software for natural language processing of Japanese sponsored by Information Promotion Agency (IPA). This engine decomposes each data sentence into words and counts the number of keyword hits against them. The total number of hits against all the words gives “similarity” of the search, and the search outputs the results in the order of higher analogy. IMAGINE (<http://imagine.bookmap.info>), developed by Takano et al. is a piece of software that put GETA together with databases of newspaper articles, encyclopedia, Wikipedia, and books [9]. All data in the databases are decomposed into sets of words in advance and recorded in the form of matrices (Fig. 2). GETA has recorded

* Corresponding author.

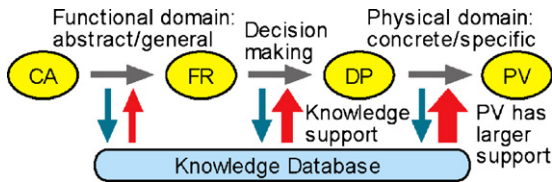


Fig. 1. Design thinking process supported by knowledge database [1].

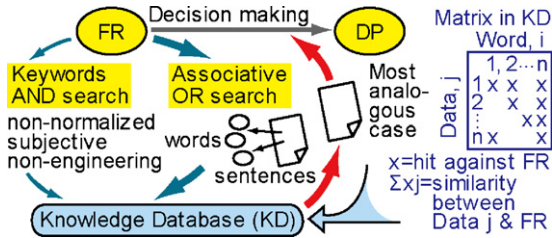


Fig. 2. Associative search for knowledge support.

matrices, for each 20 million pieces of data, which of the several million index words are used, thus calculating their inner products with the input word set vector rates the similarity, and parallel processing returns the search results within 6 s (Table 1 has not been cited in the text in the original manuscript. It has been cited in Section 2.1, para 1. Please check its citation. If not appropriate, please cite it in an appropriate place. Table 1).

2.2. Failure knowledge database

There were two databases for our study. One was "Failure Knowledge Database (FKD)" built in a project managed by Japan Science and Technology Agency (JST), published in 2006 (<http://shippai.jst.go.jp/>). This database lists 1136 accident cases from the engineering field. All cases are categorized systematically. The other was "100 Scenarios of Failure." The text of this database is the contents of a book written by one of the authors Nakao and published by Moritita Publishing Co., Ltd. in 2005. The book lists 41 failure scenario groups (Table 2) common to engineering accidents. "Scenarios" defined for our study are common and repeat in many

Table 2

The 41 scenarios in "100 Scenarios of Failure" (red #: the classified number of risk concerns, $n = 203$)

62 Technical / Dynamic scenarios	
1) 1 Brittle fracture	8) 0 Buckling
2) 12 Fatigue	9) 1 Resonance
3) 11 Corrosion	10) 1 Fluid vibration
4) 2 Stress corrosion	11) 4 Cavitation
5) 9 Polymer	12) 1 Impact
6) 11 Unbalance	13) 5 Strong wind
7) 2 Foundation collapse	14) 2 Abnormal friction
55 Technical / Anomalous	
15) 6 Non-regular use	23) 0 Brittle structure
16) 6 Fall / Attachment	24) 6 Uncontrolled feedback
17) 9 Reverse flow	25) 5 Chemical over-reaction
18) 10 Dust / Animal	26) 1 Bacteria culture
19) 1 Error accumulation	27) 2 Business complexity
20) 1 Oil flashing	28) 2 Fail-safe problem
21) 1 Fire escape	29) 3 Back-up problem
28 Technical / Operational	
30) 5 Inputting mistake	33) 3 Automation trouble
31) 8 Wiring mistake	34) 4 Use of remainders
32) 5 Piping mistake	35) 3 Adjusting in operation
18 Organizational / Social	
36) 11 Poor communication	39) 0 No change of plan
37) 7 Safety device off	40) 0 Ethical misconduct
38) 0 Illegal operation	41) 0 Terrorism
40 Non-serious / Additional to the 41 scenarios	
i) 14 Human error	iii) 9 Easy mistake of design
ii) 12 Management trouble	iv) 5 Poor planning

cases and are described in generic and abstract terms. An example is a chain reaction like, "a pump vibrates; piping resonates; weld joint fractures from fatigue; volatile liquid leaks." Of these chain reaction phenomena, a typical one (e.g., fatigue) is set to one of the scenarios in Table 2.

2.3. Seminar for extracting failure knowledge

For our experiment, Japan Society of Mechanical Engineering (JSME) and JST jointly conducted 5 sessions of the same seminar for extracting failure knowledge. A total of 90 engineers (mostly mechanical) participated in them. The participants inputted a total of 203 cases of risk concerns of their own or of their companies. The

Table 1

Examples of inputted risk concerns and outputted most analogous accidents

Similarity judgement:	Inputted risk concerns represented as product or place, as common scenario.	Outputted most analogous accidents in FKD represented as product or place, as scenario.
(a) Both agreed	A local airport had wind shear (strong traverse wind). The airplane wing hit the ground at landing.	A steel pipe factory had down burst (gust of wind). Its traveling crane ran away and crashed at the end.
	A chemical plant had rain leaking into a box-type leg. Frozen ice expanded and deformed the leg.	An excavator had snow leaking into a box-type arm. Frozen ice expanded and cracked the arm.
	A cellular phone company started a new service. Customer's overflowed access made a long-delay.	Three bank companies were merged. Customer's overflowed access made a system-down.
	Jammed papers in a shredder was pushed by a child. Fingers were pulled into the inlet and injured.	Customer snow in a blower was pushed by an old lady. Arms were pulled into the blades and injured.
(A) agreed: YES	A polymer cap was shrunk by heat in an automobile. The cap screw was loosened and leaked oil.	A polymer guide was expanded by heat in a copying machine. The guide gap widened and jammed papers.
	Motor bearings in the ceiling made high friction. Motor in a machine shop over-heated and dust fired.	Blower impellers hit its rubber guide with high friction. The over-heated blower in a chemical plant was fired.
	Vinyl chloride pipes of a tank were welded with a heat gun. The pipes were over-heated & melted.	Steel of a refrigerator was cut by acetylene gas. Its wall of urethane foam was over-heated and fired.
	An airplane had a short circuit in the fuel tank. Fuel gas fired and crashed the airplane.	Pipes beside the tank of a power plant were welded. Fuel gas in the tank fired and exploded the plant.
(b) Conflicting	The inlet rubber pipe of an oil fan heater had a crack. Less air had imperfect combustion, and some poisoned CO flowed reversely. The heater was recalled.	The safety device of a hot water heater was broken. The sensing signal of imperfect combustion was cut by a maintenance. The heater was recalled.
	The steel hub of a truck car had a fatigue fracture. The truck fell the tire and was recalled.	The work direction sensor of a crane car was broken. The crane moved reversely and was recalled.
(b') A: YES P: NO	The relay on a circuit board was switched off. Surging reflected current and broke a transistor.	Shutting flow of a water pump made water hammering. Surging reflected water and broke another pump.
	The motor of a hybrid automobile leaked current. Its bearing had electrical corrosion.	None (Later, the author found the data that the bearing of a third-rail subway had electrical corrosion.)
(c) Both disagreed	A student climbed the fence in a sightseeing bridge. He was fooling around and fell down a river.	None (Later, we found the data that a technician was fooling around in our Univ. and fell down a base.)

[Process 1] Write cases of risk concerns typed in Microsoft Excel in Japanese and copy-and-paste into the IMAGINE entry field.

Example of Excel input list in Japanese	
Incident	Accident in a local airport
Sequence of events	Wind blew a landing airplane
Cause	Strong traverse wind
Countermeasure	Doppler rader
Lesson learned	Sudden wind crashed an airplane

[Process 2] Open and read some higher-ranked analogous data on the IMAGINE output lists, and find the most analogous case as an extracted knowledge.

Example of IMAGINE output list in Japanese			
Failure Knowledge Database	Wikipedia	Newspaper articles	
rank 1 Airport	Incident	Accident in a pipe factory	
rank 2 Ocean	Sequence	Wind blew a running crane	
rank 3 Pipe factory	Cause	Strong gust of wind	
rank 4 Building	Counter	Doppler rader	
Open and read the analogous data		searched keywords	

Fig. 3. Knowledge extraction processes using associative search engine with databases, “IMAGINE”. Process 1: Write cases of risk concerns typed in Microsoft Excel in Japanese and copy-and-paste into the IMAGINE entry field. Process 2: Open and read some higher-ranked analogous data on the IMAGINE output lists, and find the most analogous case as an extracted knowledge.

input text was described in Microsoft Excel copied and pasted into IMAGINE for search by GETA as shown in Fig. 3.

Next the participants read the first five or so cases with high analogy from the search output list to find the most analogous case. This manual process was necessary because GETA weighs nouns heavier than adjectives. For example, case data frequently included abstract nouns like safety or recall, and proper nouns like city or company name; their number of appearance affected the analogy rating like noise. In other words, the “manual knowledge extraction” had to follow the “automatic information search.”

Our first experiment registered FKD to IMAGINE, finding the most analogous accident (MAA) case from it. After each session, the participants turned in a questionnaire that asked the risk concern they inputted and what the MAA found was and whether the MAA offered useful knowledge to him. Among all the 203 risk concerns, searches for 134 cases merged to MAAs within the 3.5 h seminars. Among the 134 MAAs, 113 (84%) were found from FKD, and 21 from the newspaper database because some of the risk concerns were not fully technical. The participants sometimes rated the analogy differently from the authors; we had to review the analogy rating for all the cases described in the questionnaire.

To benchmark our associative search, the authors also ran keyword searches. After the seminar was finished, we picked arbitrary 3 words from the cause description of each risk concern and ran keyword searches. If the search returned the same MAA found with associative search within the top 20 cases in the list, we judged the keyword search succeeded in locating the MAA.

For the second experiment, we registered 41 accident scenarios from the main text of “100 Scenarios of Failure” to IMAGINE and the authors of this paper ran associative search sessions for the most analogous scenario (MAS) against the above 203 risk concerns.

3. Experiment results

3.1. Results of locating MAA from Failure Knowledge Database

Fig. 4(a) shows ratio of the participants that succeeded in finding the MAA. The upper ratio of 82% is the direct results from the questionnaire as judged by the participants themselves. The

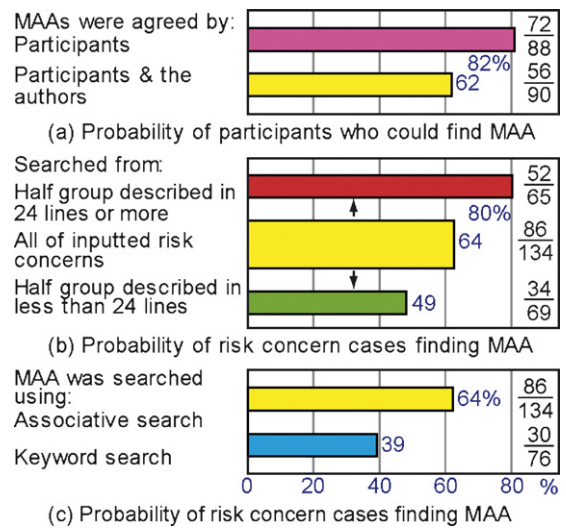


Fig. 4. Results of the seminar of locating the most analogous accident (MAA) using associative search.

lower number of 62% is the results after the authors reviewed the questionnaire and judged whether the search actually located the MAA or not. From these results, the associative search engine located analogous and useful data for reducing risk for 62% of the participants. The difference 20% of the participants, as Table 1 (band b') shows, belong to the group that gave conflicting judgments in terms of analogy of cases with the authors.

Table 1 shows examples of risk concerns and their MAAs. The yellow text shows the subject product or place and the pink the common scenario. (a) is a case where the participant and the authors agreed about the results. The subject products are different; however, the scenarios such as strong wind, frozen ice, or overflowed access are similar. (b) on the other hand, is a case where the participant claimed analogy but “recall” was the only commonality whereas the causes of accidents were different. In contrast, (b') is a case where the authors identified analogy in the scenarios of “surging” while the participant called the different product. (c) shows a case for which MAA was not identified; such risk concerns include a human error of “silly falling” or a rare case of “electrical corrosion”.

Fig. 4(b) shows the ratio of risk concern cases that successfully found useful knowledge. The middle of the figure shows the ratio 64% of all cases where the authors judged the searches were successful. The upper part shows the ratio 80% for half the cases with risk concerns described with 24 lines (equivalent to 240 words in English) or more. The lower score of 49% is for the other half with less than 24 lines. In short, for associative search to find useful knowledge, the input description has to have reasonable length with a variety of words.

To compensate for the short input description, some participants added another sentence to their OR search. For example, in searching accidents during welding, they ran an OR search with “welding” in Wikipedia; phrases like “welding robot” or “heat affected zone” were added to the input; many accident cases related to welding were found systematically.

Fig. 4(c) shows the results of keyword search. The probability of finding the same MAA was 39%. The keyword search scored about six tenths of 64% of associative search. Those of 39% that succeeded in the keyword search had causes such as ammonia corrosion, heavy snow, or bacteria; “minority” scenarios with distinct causes with 20 or less related cases in the FKD. On the other hand, those of 24% that failed the keyword search had “majority” scenarios. Of the 1136 cases in the FKD, for example, the keyword of “fire” returned 352 cases and looking for the MAA was cumbersome. Other keywords that return more than 100 cases were explosion (279), piping (251), inspection (245), corrosion (169), welding (128), and fatigue (108).

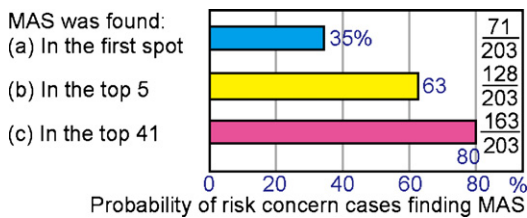


Fig. 5. Results of the seminar of locating the most analogous scenario (MAS) using associative search.

3.2. Results of locating MAS from 100 Scenarios of Failure

Fig. 5 shows the ratio of cases that successfully reached the MAS with associative search. IMAGINE, as Fig. 3 shows, lists found cases in the order of high analogy. Cases (a) with 35% are those that the MAS that the authors decided were listed in the first spot, cases (b) with 63% are those when it was listed in the top 5, and cases (c) with 80% in the top 41. The trend is the same as that described in the previous section. Most of the MAS in (a) are the “minority” cases found with the keyword search, and the result is similar to 39% for the above mentioned minority scenarios. The ratio of cases (b) is almost the same with 64% that succeeded in knowledge extraction from the first five or so cases on the list with associative search on FKD. MAS found for cases (c) are common with the technical MAA in searching FKD, not newspaper, whose ratio was 84%.

Table 2 shows the 41 scenarios and the red number of risk concerns with MAS assigned to them. Among the 230, the blue scenarios have high ratio such as fatigue (12), corrosion (11), unbalance (11), poor communication (11), and dust/animal (10). When we took a closer look at the 40 pieces of data that were not assigned to any scenario, human error (14) like traffic accident or careless mistakes, management trouble (12) like failed to follow procedure. The 100 Scenarios of Failure collected accident cases for skilled engineers and did not look at these non-serious 40 cases of individual, beginner's scenarios.

4. Discussion

Our experiment could find useful knowledge by using associative search. Our study, however, did not look at making use of the knowledge with its materialization to lower the risk after finding the knowledge.

The authors teach a junior class to design, fabricate, and test drive miniature Stirling engines. We gave an assignment of modifying their own design assuming that the product will be sold as a scientific teaching material for high-school students. The students carried out keyword search for useful knowledge from FKD and suggest modifications for safety to the engines. Among the 132 students, 108 (82%) found useful data. Even keyword search returned good results because the search area was narrow. Those, however, that turned the knowledge into good improvements

were only 63 (48%). Most failure cases upon modification were coupled designs. For example, if one, in fear of getting fingers caught in the rotation, puts a cover on the entire machine, he loses a way to start the engine by giving the initial push. Future homework for our research is the construction of an IT-assistant system that helps the designer set the DP for FR by making use of knowledge.

5. Conclusion

The process of setting the design parameter to meet FR requires knowledge described with words. This study set the FR of “reduce your own risk” and ran some experiments. 90 engineers participated in our seminar to input their own risk of 203 cases to the associative search engine “GETA/IMAGINE” to find the MAA from “Failure Knowledge Database” and the MAS from “100 Scenarios of Failure”. As a result, above 60% of the risks found MAA or MAS cases from either database in a short time of about 10 min. Keyword search, on the other hand, with the same search conditions only found MAA for only 39% of the risks. Natural language processing can support the designer in his design thinking in the functional domain, and will scientifically reveal a human creating process as the classification of abstract information, the analogy of knowledge scenarios, and the individual materialization for his own FR.

Acknowledgements

In the field of associative search, the authors are indebted to A. Takano and Y. Koike of National Institute of Informatics, H. Kinukawa and S. Kawagoe of Tokyo Denki University, and Y. Harita of Flow Net Corp. This study was sponsored by JST.

Reference

- [1] Suh NP (2001) *Axiomatic Design: Advances and Application*. Oxford University Press.
- [2] Kimura F, Ariyoshi H, Ishikawa H, Naruko Y, Yamato H (2004) Capturing Expert Knowledge for Supporting Design and Manufacturing of Injection Molds. *Annals of CIRP* 53(1):147–150.
- [3] Nakao M, Yamada S, Kuwabara M, Otubo M, Hatamura Y (2002) Decision-Based Process Design for Shortening the Lead Time for Mold Design and Production. *Annals of CIRP* 51(1):127–130.
- [4] Hon KKB, Zeiner J (2004) Knowledge Brokering for Assisting the Generation of Automotive Product Design. *Annals of CIRP* 53(1):159–162.
- [5] Chen SL, Tseng MM (2005) Defining Specification for Custom Products: A Multi-Attribute Negotiation Approach. *Annals of CIRP* 54(1):159–162.
- [6] Jin Y, Lu SC-Y (2004) Agent Based Negotiation for Collaborative Design Decision Making. *Annals of CIRP* 53(1):121–124.
- [7] Kayis B, Arndt G, Zhou M, Savci S, Khoo YB, Rispler A (2006) Risk Quantification for New Product Design and Development in a Concurrent Engineering Environment. *Annals of CIRP* 55(1):147–150.
- [8] Hatamura Y, Iino K, Tsuchiya K, Hamaguchi T (2003) Structure of Failure Knowledge Database and Case Expression. *Annals of CIRP* 52(1):97–100.
- [9] Takano A (2003) Association Computation for Information Access, Discovery Science vol.2843/2003, Springer, 33–44.